

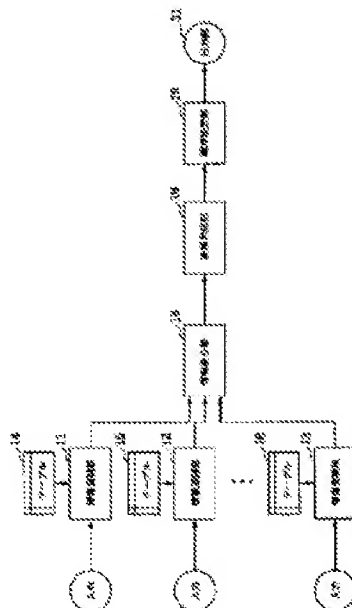
INFORMATION INTEGRATION SYSTEM, ITS METHOD AND RECORDING MEDIUM HAVING INFORMATION INTEGRATION PROGRAM RECORDED THEREON

Patent number: JP2001318947
Publication date: 2001-11-16
Inventor: MATSUSHITA MITSUNORI; KATO TSUNEAKI
Applicant: NIPPON TELEGRAPH & TELEPHONE
Classification:
- international: G06F17/30; G06F17/30; (IPC1-7): G06F17/30
- european:
Application number: JP20000139268 20000512
Priority number(s): JP20000139268 20000512

Report a data error here

Abstract of JP2001318947

PROBLEM TO BE SOLVED: To make it possible to acquire variable information also from an information source of low reliability by providing the degree of confidence of information itself simultaneously with the possibility of the information. **SOLUTION:** Information development parts 11 to 13 extend the information of information sources by referring to respective tables 14 to 16 describing the degree of the relation between the pieces of information and an information integration part 18 integrates the extended information on the basis of a similarity function of all the information sources, finds out the possibility value of information from the similarity obtained by using similarity function and finds out the reliability of information as a confidence degree value from the scatter of the similarity. An expression generation part 19 selects an expression from a possibility expression table on the basis of the possibility value of information, selects the expression from a confidence degree expression table on the basis of the confidence degree value and generates a natural sentence by the combination of the selected results. An order determination part 20 selects one of the information by applying a selection function on the basis of the possibility value and confidence degree value of each information and determines the order of the arrangement of all the information.



.....
Data supplied from the *esp@cenet* database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2001-318947
(P2001-318947A)

(43) 公開日 平成13年11月16日 (2001. 11. 16)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
G 0 6 F 17/30	3 5 0	G 0 6 F 17/30	3 5 0 C 5 B 0 7 5
	2 2 0		2 2 0 Z

審査請求 未請求 請求項の数 7 O L (全 9 頁)

(21) 出願番号 特願2000-139268 (P2000-139268)

(22) 出願日 平成12年 5 月12日 (2000. 5. 12)

(71) 出願人 000004226

日本電信電話株式会社
東京都千代田区大手町二丁目 3 番 1 号

(72) 発明者 松下 光範

東京都千代田区大手町二丁目 3 番 1 号 日
本電信電話株式会社内

(72) 発明者 加藤 恒昭

東京都千代田区大手町二丁目 3 番 1 号 日
本電信電話株式会社内

(74) 代理人 10007/274

弁理士 磯村 雅俊 (外 1 名)

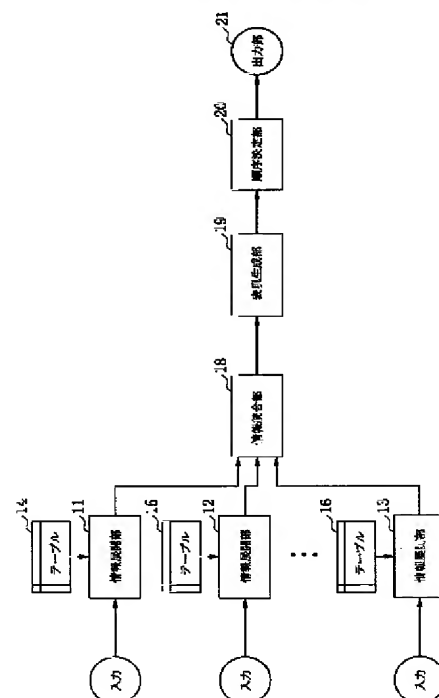
F ターム (参考) 5B075 NK54 NR15 PQ04 PR06 QM08

(54) 【発明の名称】 情報統合システムおよび情報統合方法、ならびにそのプログラムを記録した記録媒体

(57) 【要約】

【課題】 情報の可能性を提示すると同時に、情報自体の確信度を提示することができ、信頼性の低い情報源からでも価値のある情報を獲得することができる。

【解決手段】 情報展開部 1 1 ~ 1 3 は、情報源の情報を情報間の関連度を記述したテーブル 1 4 ~ 1 6 を参照して拡張し、情報統合部 1 8 は拡張した情報を全ての情報源について類似度関数に基づいて統合し、その類似度関数により得られる類似度から情報の可能性値を求めるとともに、類似度の散らばりから情報の信頼性を確信度値として求める。また、表現生成部 1 9 は、情報の可能性値に基づいて可能性表現テーブルから表現を選択し、確信度値に基づいて確信度表現テーブルから表現を選択し、組み合わせて自然文を生成する。また、順序決定部 2 0 は、各情報の可能性値と確信度値に基づいて選択関数を適用して、情報を 1 つ選択し、全体の配列順序を決定する。



【特許請求の範囲】

【請求項1】 複数の情報源の情報を比較して妥当な可能性を付与するとともに、該複数の情報を統合した情報自体の確信度を付与する情報統合システムであって、情報源の情報を情報間の関連度を記述したテーブルを参照して関連情報に展開する情報展開部と、展開した情報を全ての情報源について類似度関数に基づいて統合し、その類似度関数により得られる類似度から情報の可能性値を求めるとともに、該類似度の分散状況から情報の信頼性を確信度値として求める情報統合部と、

情報の可能性値に基づき可能性表現テーブルから表現を選択し、確信度値に基づいて確信度表現テーブルから表現を選択し、選択したものを組み合わせて自然文を生成する表現生成部と、

各情報に付与された可能性値と確信度値を基に配列順序を決定し、出力する順序決定部とを有することを特徴とする情報統合システム。

【請求項2】 複数の情報源の情報を比較して妥当な可能性を付与すると同時に、該情報を統合した情報自体の確信度を付与する方法であって、該情報源の情報を情報間の関連度を記述したテーブルを参照して関連情報に展開するステップと、展開した情報を全ての情報源について類似度関数に基づいて統合するステップと、上記類似度関数により得られる類似度から情報の可能性を求めるステップと、該類似度の分散状況から情報の信頼性を確信度値として求めるステップとを有することを特徴とする情報統合方法。

【請求項3】 請求項2に記載の情報統合方法において、前記各ステップに加えて、前記情報の可能性値に基づいて可能性表現テーブルから表現を選択するステップと、前記情報の確信度値に基づいて確信度表現テーブルから表現を選択するステップと、選択された表現を組み合わせてユーザに理解し易い自然文を生成するステップとを有することを特徴とする情報統合方法。

【請求項4】 請求項2に記載の情報統合方法において、前記各ステップに加えて、前記情報の可能性値と確信度値に基づいて選択関数を適用するステップと、該関数の値に基づいてソートするステップと、情報を1つ選択することにより、ユーザに提示する情報を決定するステップとを有することを特徴とする情報統合方法。

【請求項5】 請求項2に記載の情報統合方法において、前記類似度関数に基づいて統合するステップで用いる類

似度関数は、2つの可能性の一致度を示す関数、例えば、 w を現象、 $\Delta Poss$ を可能性の差とすると、

【数1】

$$sim(\Delta Poss_w) = e^{-10(\Delta Poss_w)^2}$$

の関数を使用することを特徴とする情報統合方法。

【請求項6】 請求項4に記載の情報統合方法において、

前記関数の値に基づいてソートするステップで用いる関数は、各情報に付与された確信度値 $Poss$ と可能性値 CF を基に次式を適用し、得られた値を降順ソートして順位を決定することを特徴とする情報統合方法。

【数2】

$$F = \frac{(\beta^2 + 1) \cdot Poss \cdot CF}{\beta^2 \cdot Poss + CF}$$

(β は、 $Poss$ 値と CF 値のどちらを重視するかを決定するパラメータ)

【請求項7】 請求項2～6に記載の情報統合方法の各ステップをプログラムに変換し、変換されたプログラムを記録媒体に格納することを特徴とするプログラム読み出し可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、可能性値と確信度値を付与して複数の情報を統合するシステムおよびその統合方法、ならびにそのプログラムを記録した記録媒体に関し、特にユーザにとり信頼性の低い情報源から価値のある情報を取得でき、かつ確信度と可能性に基づいてユーザにとって分かり易い自然言語表現を得ることが出来る情報統合システムと統合方法と記録媒体に関する。

【0002】

【従来の技術】近年、インターネットで代表されるネットワークインフラストラクチャの普及により、大量でかつ多様なデータをユーザが容易に獲得することができるようになってきた。しかし、その反面、インターネットのように情報発信の容易さを持つという性質上、情報の欠落や誤りを含む信頼性の低い情報源も氾濫するようになった。このように信頼性の低い情報源から信頼性の高い情報を抽出するためには、一般的に複数の情報源から情報を得ることが行われており、様々な情報提示システムが提案されている。従来の情報提示システムにおいて用いられる方法としては、概略以下の4種類に分類することができる。(1) 複数の情報源からの情報をそのままユーザに提示することにより、ユーザに判断を委ねる方法、(2) 情報源の間で多数決を行い、最も多くの情報源で採用されている値をユーザに提示する方法、(3) 情報源の平均値をユーザに提示する方法、(4) 最も信頼できる情報源の値を参照する方法、である。

【0003】

【発明が解決しようとする課題】しかしながら、上記（１）の方法では、ユーザが複数の情報について自分で判断しなくてはならないため、利便性の点で問題があった。また、上記（２）の方法では、例えば、情報源が３つあり、その各々が異なる情報であった場合には、異なる情報同士で多数決はとることができず、値を選択することができない、という問題があった。また、上記（３）の方法では、数値属性の場合にのみ適用できる方法であって、例えば、天候のように名義属性の情報では、その平均値を単純に得ることができない。なお、一次元の軸上にそれらの名義属性情報を配置して、値を付与することにより擬似的に数値属性化して統合することも可能ではあるが、次のような問題が依然として残る。例えば、天気予報を例にとると、晴、曇、雨の名義情報を一次元上に各々晴＝１．０、曇＝０．５、雨＝０．０として配置した場合、情報源Ａが晴、情報源Ｂが雨であったとき、その平均値は０．５となり、ユーザには曇という名義情報が提示されることになる。また、情報源Ａ、Ｂが共に曇であった場合にも、その平均値は０．５になり、ユーザには曇という名義情報が提示される。いずれも曇という予報がユーザに与えられるが、直感的には情報の信頼性が異なるにも係わらず、ユーザはそのことを知るべきがない、ということになる。特に、前者の場合には、得られた結果にも疑問が残る。さらに、前者の場合には、雨の可能性が低いながら存在するにも係わらず、統合によってその情報が失われてしまうため、例えば『雨は降るのだろうか？』というユーザの興味に適切に答えることができない、という問題点があった。

【０００４】また、上記（４）の方法では、情報源の信頼性を事前に評価しておく必要があること、また信頼性の高い情報源であったとしても、ネットワーク不調や情報源のサーバメンテナンスなどによる情報の欠如は本質的に避けられない。この点については、複数の情報源を参照する方法に比べて劣っている。また、上記（１）と（４）の方法は、情報の信頼性を評価する指標を持たないため、ユーザはその情報がどの程度信頼できるかを自分で判断しなくてはならないという問題もある。さらに、上記（２）や（３）の方法のように、情報を統合する方法では、ファジィ推論や確率推論の手法を用いて情報に可能性値を付与することはできるが、導出された信頼度がどの程度信頼できるかについては分からない。例えば、天気予報の場合、複数の情報が統合された結果、『雨の可能性が０．７である』という予測が統合情報として得られた場合、統合に用いた情報源が多いときと少ないときとでは、得られた結果自体の信頼性は自ずと異なるにも係わらず、ユーザにはそれを知るすべもない。

【０００５】上記問題を解決するために、上界確率と下界確率の２値で表現する方法として、例えば論文『推論の信頼性を考慮した不確実な知識の表現法と推論法について』（鈴木誠、松嶋敏泰、平沢茂一：情報処理学会論

文誌、Vol. 35, No.5, pp.691～705, 1994)を応用する方法があるが、この方法を上記の例に当て嵌めると『雨の可能性が０.５から０.８の範囲にある』というような表現をすることになる。この場合には、可能性の値が区間で表現されているため、統合結果の信頼性をその区間値の幅として知ることができるが、ユーザが求めているのは、どのような情報が得られるのか、そして、その情報は信頼できるのか、ということであって、区間値からユーザの求めているこれらの情報への解釈が依然としてユーザ自身に委ねられるという点で、十分な解決策とはなっていない。さらに、同じ尺度の複数の数値（この場合には二つの可能性値）が情報に付与されると、ユーザの直感的理解を損ねることにもなってしまう。

【０００６】そこで、本発明の目的は、これら従来の課題を解決し、情報の可能性を提示するだけでなく、統合された情報がどの程度信頼できるかを提示でき、ユーザにとって信頼性の低い情報源からでも価値のある情報を獲得でき、また確信度と可能性に基づいてユーザにとって分かり易い自然言語表現を得ることができる情報統合システムおよび情報統合方法を提供することである。

【０００７】

【課題を解決するための手段】上記目的を達成するため、本発明の情報統合システムでは、①複数の情報源の情報を比較して妥当な可能性を付与すると同時に、その統合した情報自体の確信度を付与するシステムであって、情報源の情報を情報間の関連度を記述したテーブルを参照して関連情報に展開し、展開した情報を全ての情報源について類似度関数に基づいて統合し、その類似度関数により得られる類似度から情報の可能性値を求めると同時に、その類似度の散らばりから情報の信頼度値として求めるようにしている。また、②上記①のシステムにより得られる情報の可能性値と確信度値に基づいて、ユーザに理解し易い自然文を生成する方法であって、該情報の可能性値に基づいて可能性表現テーブルから表現を選択し、確信度値に基づいて確信度表現テーブルから表現を選択し、組み合わせる自然文を生成するようにしている。さらに、③上記①のシステムにおいて得られる複数情報の可能性値と確信度値に基づいてユーザに提示する情報を決定する方法であって、各情報の可能性値と確信度値に基づいて選択関数を適用して情報を１つ選択するものである。

【０００８】本発明においては、上記①のシステムにより、複数の情報源の情報を比較して妥当な可能性を付与すると同時に、その統合した情報自体の確信度を付与することができるので、統合された情報がどの程度信頼できるかを直感的に理解できることから、分かり易いという利点がある。また、上記②の方法により、該情報の可能性値に基づいて可能性表現テーブルから表現を選択し、確信度値に基づいて確信度表現テーブルから表現を選択し、組み合わせる自然文を生成するようにしている

ため、得られる情報の可能性値と確信度値に基づいてユーザに理解し易い自然文を生成することができるという利点がある。さらに、上記㉓の方法により、得られる複数情報の可能性値と確信度値に基づいてユーザに提示する情報を決定することができるという利点がある。

【0009】

【発明の実施の形態】以下、本発明の実施例を、図面により詳細に説明する。図1は、本発明の一実施例を示す情報統合システムのブロック構成図である。本発明の情報統合システムは、複数の情報源からの情報を統合して可能性値と確信度値を伴う複数の情報を出力すると共に、それらの情報に対する自然文表現を生成する。図1において、ある情報ソース（入力）から得られた情報を関連度に基づいて可能性値を付与して展開する情報展開部11～13と、情報源の情報を情報間の関連度を記述したテーブル14～16と、展開された複数の情報源からの情報を統合し、確信度を付与する情報統合部18と、付与された可能性値と確信度値を基に自然言語表現を生成する表現生成部19と、該可能性値および確信度値を基に順位を決定する順序決定部20とから構成されている。ここで、もし自然言語表現をする必要がなければ、表現生成部19の処理を行う必要はない。また、順序付けする必要がなければ、順序決定部20の処理を行う必要はない。

【0010】図2は、図1における情報展開部の処理フローチャートである。なお、図6は情報関連テーブルのデータ構成例を示し、図7は展開情報テーブルのデータ構成例を示している。具体例として、天気予報情報の統合を例にとって説明する。図1の各入力には、個々の情報源からの情報が入力される。天気予報の例において、情報源とは天気情報を提供するWEBサイトなどを指している。また、情報とは、『晴』や『曇』などの天気予報を指している。本発明では、各情報源から入力される情報を受け付け（ステップ101）、各々の情報源に対応する情報展開部11～13に送られる。情報展開部11～13は、図6に示すような情報関連テーブル10を参照し（ステップ102）、入力された情報を展開する。なお、このテーブル10は、図6に示すように、観測値と各情報源の差分から導出もしくは統合した結果と、各情報源の差分からの導出もしくは人手による作成などの様々な方法で作成することができる。ここでは、その作成方法については言及しない。

【0011】情報展開部11～13は、入力を与えられたならば、図6に示す情報関連テーブル10を参照して、入力と一致する入力天気を持つ行を取り出し（ステップ103）、全ての一致した行を出力して（ステップ104）、情報統合部18に渡して処理を終了する。例えば、情報源Aから情報展開部11～13に与えられた入力『晴』であった場合には、図6の情報関連テーブル10を参照して、入力と一致する関連情報テーブルの

入力天気の行を選択することにより、図7に示すような展開情報テーブルが得られることになる。図7に示すように展開された情報は、全て情報統合部18に送出される。全ての情報源について、同様の処理が行われる。ただし、常に全ての情報源に情報が存在する必要はない。

【0012】図3は、図1における情報統合部の処理フローチャートである。なお、図8は信頼係数テーブルのデータ構成例、図9および図10は情報源BおよびDからの各展開情報テーブルのデータ構成例、図11は各情報源の展開情報を合成したテーブルのデータ構成例、図12は図11の内容を補完したテーブルのデータ構成例、図13は情報源の信頼係数のテーブルデータ構成例、図14は統合テーブルのデータ構成例である。情報統合部18は、各情報源の情報展開部11～13から展開情報テーブルを受け取り、同じ情報ごとに統合し、各々の情報に可能性値と確信度値を付与する処理を行う。ここで、可能性値とは、その情報が真である可能性を示す値であり、確信度とは、その可能性値がどれだけ信頼の置けるものであるのかを示す値である。すなわち、情報統合部18は、まず展開情報テーブルを読み込み（ステップ201）、情報が存在する情報源と存在しない情報源を識別し、存在しない情報源を計数する（ステップ202）。情報が存在する情報源（これを統合対象情報源と呼ぶ）の数に基づいて、図8に示すような信頼係数テーブル10cを参照して、統合情報の信頼係数を決定する（ステップ203）。例えば、情報源がA、B、C、Dの4つあり、情報源Cから情報が獲得できなかった場合、統合対象情報源は3であるため、信頼係数は図8の信頼係数テーブル10cから信頼係数0.9と決定される。

【0013】次に、統合対象情報源から受け取った展開情報テーブルを参照し、含まれている情報の和集合を生成する（ステップ204）。そして、この生成された情報の和集合と統合対象情報源からなる2次元テーブルを作成する（ステップ205）。例えば、情報源がA、B、C、Dと4つあり、情報源Cから情報が獲得できなかった場合に、情報源Aからの展開情報テーブルが図7の展開情報テーブルであり、情報源Bからの展開情報テーブルが図9に示す展開情報テーブル、情報源Dからの展開情報テーブルが図10に示す展開情報テーブルとする。次に、これらのテーブルから図11に示す合成テーブルが作成される。そして、各統合対象情報源の値を上記合成テーブルのセルに入力する（ステップ206）。次に、図11の合成された展開情報テーブルに対して、値のないセルに0.0を補完する（ステップ207）。この処理により、図12に示す補完テーブルが得られる。次に、上記処理により補完された図12のテーブルに基づいて情報の統合を開始し（ステップ208）、確信度と可能性を決定する。なお、ステップ208とステップ213とは同じ処理の開始ブロックと終端ブロック

を意味しており、時間的にこの間継続して情報を統合する処理を行っていることを表わしている。次に、各可能性値に対して確信度を計算する（ステップ209）。

【0014】情報の統合は、次のようにして行う。すなわち、統合対象の情報源集合を $S = \{s_1, \dots, s_j\}$ とする。ある情報源 $s_j (\in S)$ において、ある

$$CF(Poss_w(s_j)) = \frac{\sum_{k=1}^J \alpha(s_k) \cdot \text{sim}(Poss_w(s_j) - Poss_w(s_k))}{\sum_{k=1}^J \alpha(s_k)}$$

ここで、関数 $\text{sim}()$ は、二つの可能性の一致度を示す関数である。 $\text{sim}()$ は、様々な関数が適用可能である。例えば、下式（数4）の関数を使用することができる。この関数は、類似度として各情報源から得られる可能性の類似を意味し、 $\text{sim}(\Delta Poss_w)$ が大きければ類似度は高く、 $\text{sim}(\Delta Poss_w)$ が小さければ類似度は低いと考えられる。

【数4】

$$\text{sim}(\Delta Poss_w) = e^{-10(\Delta Poss_w)^2}$$

$$\overline{Poss_w(s)} = \arg \max_{s_j \in S} CF(Poss_w(s_j))$$

$\overline{Poss_w(s)}$ は CF 値を最大にする $Poss_w(s_j)$ を意味する。ただし、複数の $Poss_w(s_j)$ で同じ確信度をとる場合は可能性の大きいものを採択する。また確信度は、 $\overline{Poss_w(s)}$ に対応する CF 値に上記で求めた信頼係数を掛けた値を採択する。

このようにして、確信度が計算で求められたならば、確信度の最も大きい可能性値を統合情報の可能性値に採用して、統合テーブルに書き込む（ステップ210）。上記可能性値の確信度を統合情報の確信度として採用し（ステップ211）、上記確信度値に信頼係数を掛けて、統合テーブルに書き込む（ステップ212）。これらの処理により、図14に示すような統合テーブルが生成される。統合された情報のリストは、表現生成部19に渡すために出力される（ステップ214）。

【0016】図4は、図1における表現生成部の処理フローチャートである。なお、図15は可能性値対応テーブルのデータ構成例、図16は確信度値対応テーブルのデータ構成例である。表現生成部19は、可能性値と確信度値に基づいて、対応する言語表現を生成する。ただし、表現の生成が不要な場合には、この処理は行わなくてもよい。まず、表現生成部19は、情報統合部18から渡された統合テーブルを読み込み（ステップ301）、読み込んだテーブルの各行の取り出しを開始する（ステップ302）。なお、なお、ステップ302とステップ307とは同じ処理の開始ブロックと終端ブロックを意味しており、時間的にこの間継続してテーブルの各行の取り出し処理を行っていることを表わしている。次に、表現生成部19は、先ず各情報の可能性値に基づいて言語表現を生成する。図15は、可能性値と表現との対応関係を示したテーブル10aである。ただし、0

気象事象 w が起こる可能性は $Poss_w(s_j)$ で与えられるものとする。気象事象 w の起こる可能性が $Poss_w(s_j)$ である確信度 $CF(Poss_w(s_j))$ は、下式（数3）により求められる。

【数3】

また、 $\alpha(s)$ は情報源 s の信頼係数を意味する重みであって、 $[0, 1]$ の値をとる。つまり、 $\alpha(s)$ が1の時に最も信頼度が高く、0の時に最も信頼度が低い。

【0015】次に、図13に示すような情報源の信頼係数テーブルを参照して、 $\alpha(s)$ を決定する。 $\alpha(s)$ は静的である必要はなく、更新頻度や最新更新日時、情報源の母数などに応じて変化しても差し支えない。統合した際の情報 w の $Poss$ 値（可能性値）は、下式（数5）を満たすものを採用する。

【数5】

$a < b < c < 1$ で、〔情報〕には統合テーブルの情報の項目の値（名義属性天気予測、例えば『晴』など）が挿入される。可能性値対応テーブル10aを参照することにより、可能性値に対応する表現を可能性値対応テーブル10aより取り出す（ステップ303）。

【0017】次に、表現生成部19は、各情報の確信度値に基づいて言語表現を生成する。図16は、確信度値と表現との対応関係を示したテーブル10bである。この確信度値対応テーブル10bを参照することにより、確信度値に対応する表現を確信度値対応テーブル10bより取り出す（ステップ304）。ただし、 $0 < d < e < f < g < 1$ で、〔可能性〕には図16のテーブル10bの可能性値によって決定された表現が挿入される。例えば、 $a = 0.2$, $b = 0.6$, $c = 0.8$, $d = 0.6$, $f = 0.8$ とすると、上記処理により、図17に示す表現付統合テーブルが得られる。このようにして、上記ステップで得られた表現を合成する（ステップ305）。この処理によって、生成された表現を統合テーブルに追加する（ステップ306）。表現付統合テーブルを順序決定部20に渡すために出力する（ステップ308）。

【0018】図5は、図1における順序決定部の処理フローチャートである。なお、図18は、ソートされた表現付統合テーブルのデータ構成例を示す図である。順序決定部20は、表現生成部19から渡された統合テーブ

ルの読み込みを行い（ステップ401）、順序決定部20は、順序算出関数を各情報の可能性値と確信度値に適用する（ステップ402）。すなわち、順序決定部20は、各情報に付与された確信度値と可能性値を基に下式（数6）を適用し、得られた値を降順にソートして順位を決定する（ステップ403）。

【数6】

$$F = \frac{(\beta^2 + 1) \cdot Poss \cdot CF}{\beta^2 \cdot Poss + CF}$$

ここで、Poss値は可能性値、CF値は確信度値を意味する。また、 β はPoss値とCF値のどちらを重視するかを決定するパラメータであり、ユーザもしくはシステムが事前に設定する。 β が1のとき、調和平均（ n 個の数 x_1, x_2, \dots, x_n に対して $1/x_1, 1/x_2, \dots, 1/x_n$ のように、相加平均の逆数）を意味する。例えば、図17の統合テーブルが入力されたとき、 $\beta=1$ の場合、このテーブルはソートされて図18のテーブルが得られる。これは、可能性値の降順にソートされている。ソート結果（図18のテーブル）を出力部21から出力する（ステップ404）。

【0019】このようにして、図17のテーブルのように、複数の表現が可能性値と確信度値を伴って生成された後に、これらを順序付けるために式〔数4〕のF値を用いる。各行でF値を計算し、Fの値が大きい行から順に並べ換える。例えば、『明日の天気は？』と尋ねられたとき、全ての情報を提示するのは対話システムとして不適当であるため、最も適切な文の1つを発話するようにするためにソートする。図18のテーブルがソートされた結果であり、このテーブルから、最もF値が高かった、つまり最も上位にある『確実に晴れである』が発話されるようになる。なお、ここでは、1つを選択しているが、上位3つのように複数でも差し支えない。

【0020】図2～図5のフローチャートの各ステップをプログラムに変換し、変換されたプログラムをCD-ROM、ハードディスク等の記録媒体に格納しておくことで、任意のパーソナルコンピュータにこの記録媒体を装着し、記録媒体から主メモリにインストールするか、ネットワークを介して他の任意のパーソナルコンピュータの主メモリにダウンロードすることにより、それらのパソコンでプログラムを実行すれば、任意の場所で本発明の情報統合方法を実現することができる。

【0021】

【発明の効果】以上説明したように、本発明によれば、複数の情報源の情報を比較して妥当な可能性を付与するとともに、その統合した情報自体の確信度を付与するこ

とにより、情報の可能性を提示すると同時に統合された情報がどの程度信頼できるかを提示することが可能になる。従って、ユーザにとって信頼性の低い情報源からでも価値のある情報を獲得することができる。また、確信度と可能性に基づいて、ユーザにとって分かり易い自然言語表現を取得することが可能となる。

【図面の簡単な説明】

【図1】本発明の一実施例を示す情報統合システムの全体構成図である。

【図2】図1における情報展開部の処理フローチャートである。

【図3】図1における情報統合部の処理フローチャートである。

【図4】図1における表現生成部の処理フローチャートである。

【図5】図1における順序決定部の処理フローチャートである。

【図6】本発明で使用される情報関連テーブルのデータ構成例を示す図である。

【図7】本発明で使用される展開情報テーブルのデータ構成例を示す図である。

【図8】本発明で使用される信頼係数テーブルのデータ構成例を示す図である。

【図9】本発明で使用される情報源Bからの展開情報テーブルのデータ構成例を示す図である。

【図10】本発明で使用される情報源Dからの展開情報テーブルのデータ構成例を示す図である。

【図11】本発明で使用される各情報源の展開情報を合成したテーブルのデータ構成例の図である。

【図12】図11を補完したテーブルのデータ構成例を示す図である。

【図13】本発明で使用される情報源の信頼係数のテーブルを示す図である。

【図14】本発明で使用される統合テーブルの図である。

【図15】本発明で使用される可能性値対応テーブルの図である。

【図16】本発明で使用される確信度値対応テーブルの図である。

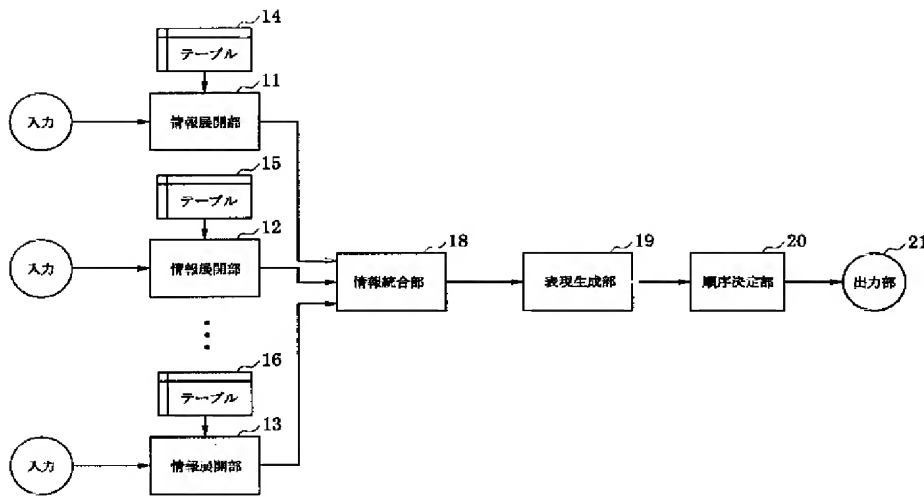
【図17】本発明で使用される表現付統合テーブルの図である。

【図18】本発明で使用されるソートされる表現付統合テーブルの図である。

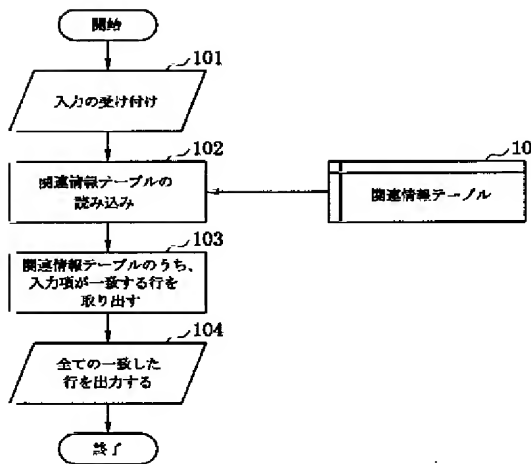
【符号の説明】

11～13…情報展開部、14～16…情報関連テーブル、18…情報統合部、19…表現生成部、20…順序決定部、21…出力部。

【図1】



【図2】



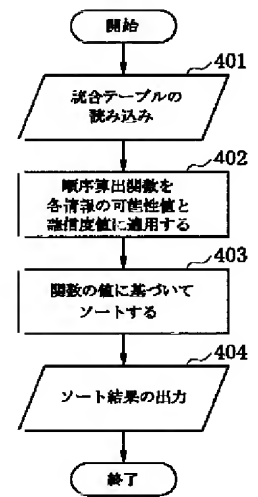
【図6】

情報関連テーブル		
入力天気	関連天気	可能性
晴	晴	1.0
晴	晴時々曇	0.8
晴	晴時々雨	0.6
晴	曇	0.6
晴	曇時々晴	0.3
晴	雨	0.1
晴	雨時々晴	0.4
曇	曇	1.0
曇	雨時々曇	0.6
...
雪後雨	雪	0.8

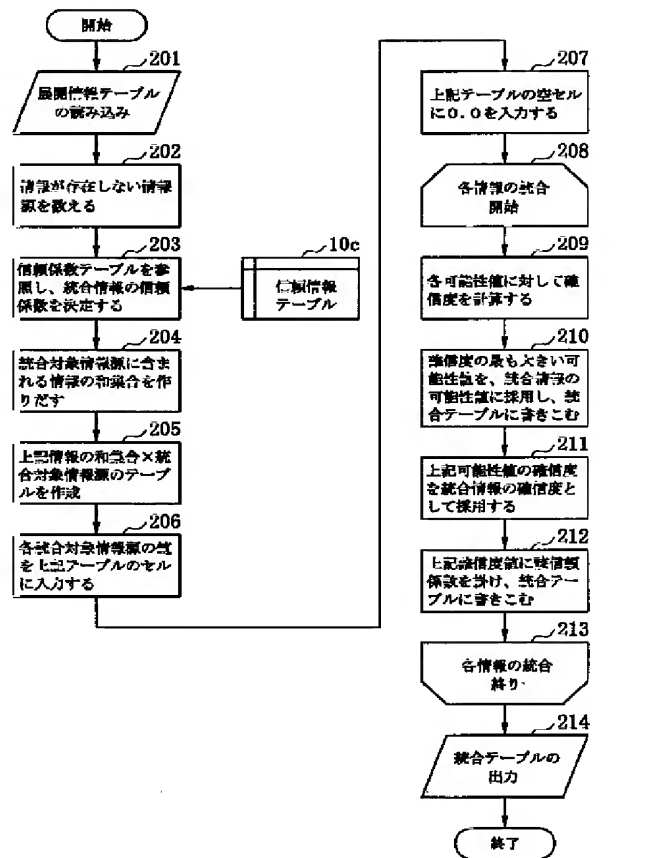
【図7】

関連情報テーブル	
関連天気	可能性
晴	1.0
晴時々曇	0.8
晴時々雨	0.6
曇	0.6
曇時々晴	0.3
雨	0.1
雨時々晴	0.4

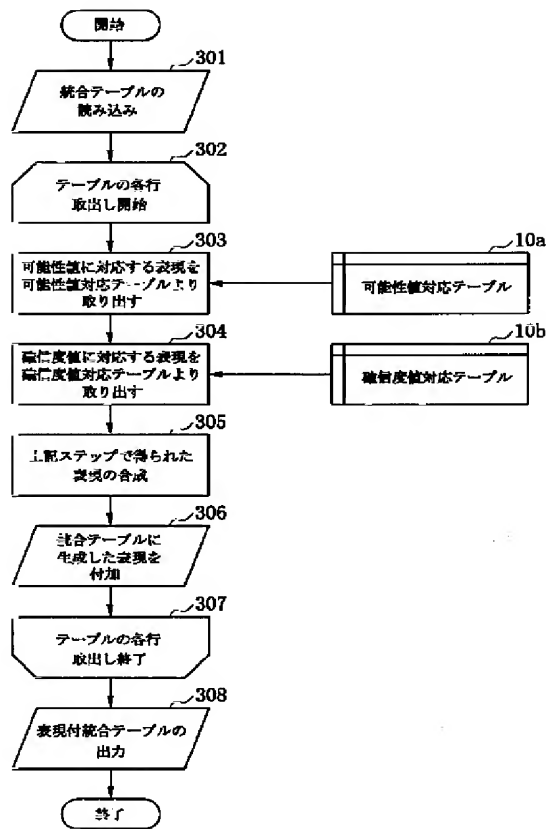
【図5】



【図3】



【図4】



【図11】

各情報源の展開情報を合成したテーブル

情報	情報源A	情報源B	情報源D
晴	1.0	0.7	0.8
晴時々曇	0.8	0.5	1.0
晴時々雨	0.6	1.0	0.6
晴後曇	0.5		0.8
晴後雨	0.6	0.9	0.4
曇	0.6		0.6
曇後晴	0.8	0.3	0.1
曇後雨		0.4	
雨	0.1	0.6	
雨時々晴	0.6		
晴一時雨			0.6

【図15】

可能性値対応テーブル

可能性値	表現
$0 \leq Poss < a$	【情報】ではない。
$a \leq Poss < b$	【情報】である可能性は低い。
$b \leq Poss < c$	【情報】である可能性が高い。
$c \leq Poss \leq 1$	【情報】である。

【図8】

信頼係数テーブル

情報源数	係数
4以上	1
3	0.9
2	0.8
1	0.7

【図12】

図11を補充したテーブル

情報	情報源A	情報源B	情報源D
晴	1.0	0.7	0.8
晴時々曇	0.8	0.5	1.0
晴時々雨	0.6	1.0	0.6
晴後曇	0.5	0.0	0.8
晴後雨	0.6	0.9	0.4
曇	0.6	0.0	0.6
曇後晴	0.8	0.3	0.1
曇後雨	0.0	0.4	0.0
雨	0.1	0.6	0.0
雨時々晴	0.6	0.0	0.0
晴一時雨	0.0	0.0	0.6

【図16】

確信度値対応テーブル

確信度	表現
$0 \leq CF < d$	【可能性】とおもうのですが、自信がありません。
$d \leq CF < e$	多分【可能性】だと思います。
$e \leq CF < f$	【可能性】と思います。
$f \leq CF \leq 1$	確実に【可能性】。

【図9】

情報源Bからの展開情報テーブル

天気	可能性値
晴	0.7
晴時々曇	0.5
晴時々雨	1.0
晴後雨	0.9
曇後晴	0.3
曇後雨	0.4
雨	0.6

【図13】

情報源の信頼係数

情報源s	信頼係数 $\alpha(s)$
情報源A	1.0
情報源B	0.9
情報源C	0.7
情報源D	0.9

【図14】

統合テーブル

情報	可能性	確信度
晴	0.8	0.85
晴時々曇	0.8	0.70
晴時々雨	0.6	0.74
晴後曇	0.5	0.51
晴後雨	0.6	0.70
曇	0.6	0.69
曇後晴	0.3	0.57
曇後雨	0.4	0.71
雨	0.1	0.67
雨時々晴	0.0	0.65
晴一時雨	0.0	0.69

【図17】

表現付統合テーブル

情報	可能性	確信度	表現
晴	0.8	0.85	確実に晴である。
晴時々曇	0.8	0.70	晴時々曇であると思います。
晴時々雨	0.6	0.74	晴時々雨である可能性が高いと思います。
晴後曇	0.5	0.51	晴後曇である可能性は低いとおもうのですが、自信がありません。
晴後雨	0.6	0.70	晴後雨である可能性が高いと思います。
曇	0.6	0.69	多分曇である可能性が高いと思いますが。
曇後晴	0.3	0.57	曇後晴である可能性は低いとおもうのですが、自信がありません。
曇後雨	0.4	0.71	曇後雨である可能性は低いと思います。
雨	0.1	0.67	多分雨ではないと思いますが。
雨時々晴	0.0	0.65	多分雨時々晴ではないと思いますが。
晴一時雨	0.0	0.69	多分晴一時雨ではないと思いますが。

【図18】

ソートされた表現付統合テーブル

情報	可能性	確信度	表現
晴	0.8	0.85	確実に晴である。
晴時々曇	0.8	0.70	晴時々曇であると思います。
晴時々雨	0.6	0.74	晴時々雨である可能性が高いと思います。
晴後雨	0.6	0.70	晴後雨である可能性が高いと思います。
曇	0.6	0.69	多分曇である可能性が高いと思いますが。
曇後雨	0.4	0.71	曇後雨である可能性は低いと思います。
晴後曇	0.5	0.51	晴後曇である可能性は低いとおもうのですが、自信がありません。
曇後晴	0.3	0.57	曇後晴である可能性は低いとおもうのですが、自信がありません。
雨	0.1	0.67	多分雨ではないと思いますが。
雨時々晴	0.0	0.65	多分雨時々晴ではないと思いますが。
晴一時雨	0.0	0.69	多分晴一時雨ではないと思いますが。